

Linking Health Records: Human Rights Concerns*

Fritz Scheuren, Ernst and Young, LLP

1. Purpose

The purpose of this paper is to provide an introduction or "starter set" for reflecting on human rights issues that arise when bringing together or linking the health records of individuals. In particular, the paper will discuss the potential role of record linkages in the proposed new United States health information system; specifically, how linkage applications may affect both the rights of individuals to privacy and their rights of access to health care services.

Four potential types of record linkages will be covered (see Figure 1 below). The primary concern will be with linkages of health records, such as the computerized enrollment and encounter records proposed to be created under the Health Security Act or other health care reform legislation [1]. As the columns of Figure 1 indicate, linkages for both statistical and administrative purposes will be considered. As the rows of Figure 1 imply, there will be a discussion of record linkage within the health system, e.g., records of individuals may be linked to records of providers or insurers. The paper will also consider linkages of health care records with records from other systems, such as vital records or social security, income tax, and welfare program records.

In all, the paper is organized into eight sections: the present introduction and statement of purpose (Section 1); a background section on what is meant by record linkage -- both in general and with respect to health record systems (Section 2); then there are four short sections, each devoted to a cell in Figure 1 (Sections 3 to 6); and, finally, a brief overall summary with recommendations (Section 7). The main questions to be addressed throughout are the extent to which linkages should be permitted, for what purposes, and under what conditions. An Afterword has been included (as Section 8) to afford room for a more personal comment.

Figure 1. -- Potential Types of Health Record Linkages
(Cell entries reference paper section where topic covered)

Linkages	Purposes	
	Administrative	Statistical
Within health record system	Section 3	Section 4
With other Record systems	Section 6	Section 5

*Reprinted with permission. See Note at end of paper.

2. Background

This section is a review of automated record linkage techniques, the nature of record linkage errors, and some overall system concerns in a world where multiple opportunities exist to carry out record linkages.

2.1 Types of Record Linkages

It seems fairly safe to speculate that once human beings began to keep records there were efforts to link them together. Until well into this century, though, such work was done manually and often only with great difficulty and expense; however, there now exist four broad types of automated record linkage (see Figure 2) -- each of which will be described below by means of an example.

Figure 2. -- Examples of Linkage Types and System Structures

Type of Record Linkage	Record System Structure	
	Intended for Linkage	Incidental to Linkage
Deterministic	Social Security and Medicare systems	National Death Index (NDI)
Probabilistic	1990 Census Post Enumeration Survey	NDI Links to the Current Population Survey

In the United States, the first National experience with automated record linkage systems was the assignment, beginning in 1935, of social security numbers (SSN's) to most wage workers. Initially this system was based on a single punch card for each worker; these cards were updated using the SSN as an account identifier and a cumulative total kept of taxable wages received under covered employment. Record linkages at the Social Security Administration were computerized in the 1950's and SSN's are issued now to virtually all Americans.

From its inception, the intended use of the social security number was to carry out record linkage. Efforts, not always successful, were made so that SSN's, when assigned, would be unique and each person would have just one [2]. Further, the wage reporting system was designed so that updates by SSN would be conducted in a manner relatively free of error. Put another way, the social security system was designed or **intended** all along for automated record linkage and a straightforward, so-called **deterministic** linkage rule of **exact matching** on SSN's was to be the basic approach.

Birth and death registration in the U.S. offers a useful contrast to social security. These vital registers, which became complete only in the 1930's, were not intended for automated linkage operations [3]. Identifying items, like names, are on these records, of course, and could be used as matching keys but would not always be unique alone -- common surnames like Smith or Johnson or Williams being notable cases where linkage problems might be particularly severe. Automated linkages to U.S. death records did not begin nationally until the inception in the 1970's of the National Death Index or NDI. The NDI in its original operations relied

on **multiple exact matches** as a way to locate potential linkages; [4] hence, as shown in Figure 2, the NDI may serve as an example of a **deterministic** automated linkage approach that was **added on** to a system not initially designed for such a use.

Deterministic match rules are easy to automate but do not adequately reflect the uncertainty that may exist for some potential links. They can also require costly manual intervention when errors occur in the matching keys. More complicated methods were needed that weighed the linkage information, allowing for errors and incompleteness, and minimizing the clerical intervention required to select the best link from all those possible. Such techniques are called **probabilistic**. The main theoretical underpinnings for probabilistic matching methods were firmly established by the late nineteen sixties with the papers of Tepping [5] and, especially, Fellegi and Sunter [6]. Sound practice dates back even earlier, at least to the nineteen fifties and the work of Newcombe and his collaborators [7].

The Fellegi-Sunter approach is basically a direct extension of the classical theory of hypothesis testing to the problem of record linkage. A mathematical model is developed for recognizing records in two files which represent identical units (said to be matched). As part of the process there is a comparison between all possible pairs of records (one from each file) and a decision made as to whether or not the members of the comparison-pair represent the same units, or whether there is insufficient evidence to justify either of these decisions. The three outcomes from this process can be referred to as a “link,” “nonlink,” or “potential link.”

In point of fact, Fellegi and Sunter contributed the underlying theory to the methods already being used by Newcombe and showed how to develop and optimally employ probability weights to the results of the comparisons made. They also dealt with the implications of restricting the comparison pairs to be looked at, that is of “blocking” the files, something that generally has had to be done when linking files that are at all large.

Many of the major public health research advances made in recent decades have benefitted at least in part from probabilistic linkage techniques. Included are such well known epidemiological findings as the effects of smoking, risks from radiation exposure, asbestos and many other carcinogens arising in the workplace, through diet or other exposures -- increasingly in populations with genetic predispositions [8]. These benefits have to be considered when exploring record linkage impacts on privacy and other rights. We will return to this point at the end of this paper where trade-offs are explicitly considered.

Most of these automated linkages, like Newcombe’s studies of radiation exposure at Chalk River (and elsewhere), were not envisioned when the records were originally created. Some probabilistic linkage systems were intended, however -- notably for “post enumeration” surveys (PES’s), carried out to evaluate U.S. decennial census coverage. For example, the PES for 1990 was particularly well designed for carrying out probabilistic linkages [9]. Another good example of a continuing probabilistic linkage that has been a real success for statistical purposes is the bringing together of the NDI and Current Population Survey [10]. This linkage, though, was not planned into the design of either of the data sets being employed.

2.2 Nature of Linkage Errors and Identifying Information

All linkage operations are subject to two main types of errors: matching records together that belong to different entities (false matches) and failing to put records together that belong to the same entity (false non-matches). These errors can have different human rights implications, depending on what the linkages are used for (see Figure 3).

Figure 3. -- Linkage Error Implications on Human Rights

	Types of Linkage Error	Linkages Used for --	
406 ■			

False Matches	Data about that individual	Information about a class of individuals
False Nonmatches	Potentially very serious	May be less serious

If the linkage is to assemble data about an individual so an administrative or diagnostic determination can be made about *that* individual, then the consequences of any error could be grave indeed. Potentially, a different (lower) standard of accuracy could be tolerated, provided a suitable adjustment is made when analyzing the results of linkage operations whose purpose is to obtain information about a group [11]. More will be said about these issues in later sections, particularly how this distinction affords an opportunity to both preserve individual privacy rights -- through group matches, say -- but still attain societal information needs.

If an efficient (low cost, essentially error free) health care linkage system is a goal, then consideration needs to be given to the establishment of a health identification "number." In ideal circumstances, personal identifying information on a medical record should satisfy the following requirements [12].

- The identifying information should be **permanent**; that is, it should exist at the birth of a person to whom it relates or be allocated to him/her at birth, and it should remain unchanged throughout life.
- The identifying information should be **universal**; that is, similar information should exist for every member of the population.
- The identifying information should be **reasonable**; that is, the person to whom it relates and others, should have no objection to its disclosure for medical purposes.
- The identifying information should be **economical**; that is, it should not consist of more alphabetic, digits and other characters than necessary.
- The identifying information should be **simple**; that is, it should be capable of being handled easily by a clerk and computers.
- The identifying information should be **available**.
- The identifying information should be **known**; that is, either the person to whom it relates or an informant acting on his/her behalf should be able to provide it on demand.
- The identifying information should be **accurate**; that is, it should not contain errors that could result in its discrepancy on two records relating to the same person.
- The identifying information should be **unique**; that is, each member of the population should be identified differently.

The social security number, incidentally, fails several of these tests. Only now is it beginning to be issued at birth; also it is far from being accurately reported. In practice, too, because of incentives created by the SSN's use in the tax system, the number is not always unique. Some people use *more than one* SSN, even in

the same year, and more often over longer periods of time. Multiple uses of the *same* SSN by different people have been common , as well.

Concerns about the risks to health records from unauthorized disclosures are greater with an identifier like the SSN which is widely available on many large private data bases, like credit files, and of course many non-health related Federal, state and other government files [13]. In the Office of Technology Assessment's 1993 report [14] on privacy the following recommendation is made with regard to the SSN.

The use of the social security number as a unique patient identifier has far-reaching ramifications for individual health care information privacy that should be carefully considered before it is used for that purpose.

Elsewhere [15] the stronger recommendation has been made not to use the SSN as a health identifier. Its use could lead to matching errors and might greatly increase the potential for unregulated linkages between health and nonhealth data sets.

2.3 Some Proposed Health Record Linkage Systems

The proposed Health Security Act [16] calls for the establishment of a National Health Board to oversee the creation of an electronic data network. The types of information collected would include: enrollment and disenrollment in health plans; clinical encounters and other items and services from health care providers; administrative and financial transactions and activities of participating states, regional alliances, corporate alliances, health plans, health care providers, employers, and individuals; number and demographic characteristics of eligible individuals residing in each alliance area; payment of benefits; utilization management; quality management; grievances, and fraud or misrepresentation in claims or benefits [17].

The Health Security Act specifies, among other things, the use of uniform paper forms containing standard data elements, definitions, and instructions for completion; requirements for use of uniform health data sets with common definitions to standardize the collection and transmission of data in electronic form; uniform presentation requirements for data in electronic form; and electronic data interchange requirements for the exchange of data among automated health information systems.

A prototype health care record linkage system may be worth considering as well since it spells out an initial schematic of a person-level health or patient record. Data could come from an array of health care settings, linked together using a "linkage processor." This processor would determine the linkage and also assign the unique patient identifier in the actual patient record. Record types would differ by the type of provider from which they are derived. The functions of the record linkage software program are outlined in Figure 4. It is anticipated that the patient identifying information would be housed in a person's primary care unit. The linkage processor stores the patient identifying data and generates the unique identifier. It processes records from other providers and links the record as shown. Some initial data categories and identifying information are outlined in Figure 5 [18].

Figure 4. -- Patient Record Prototype

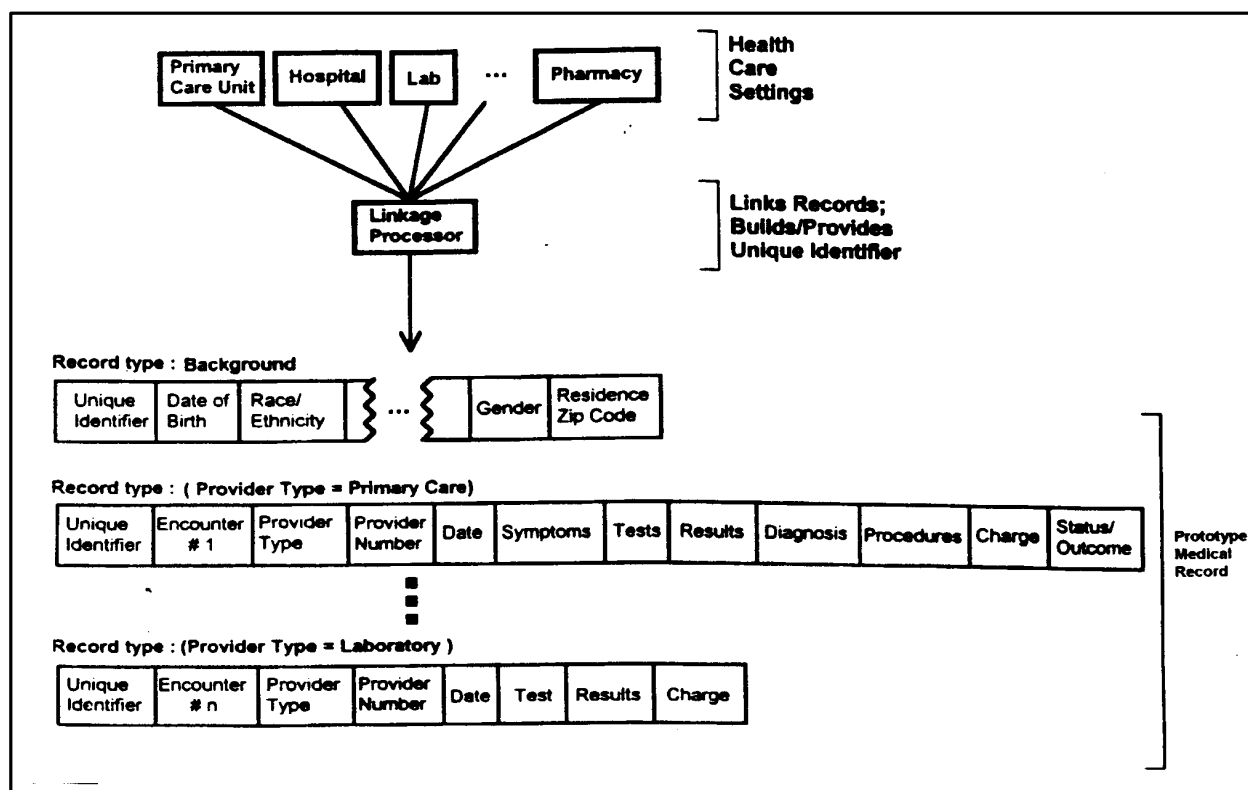
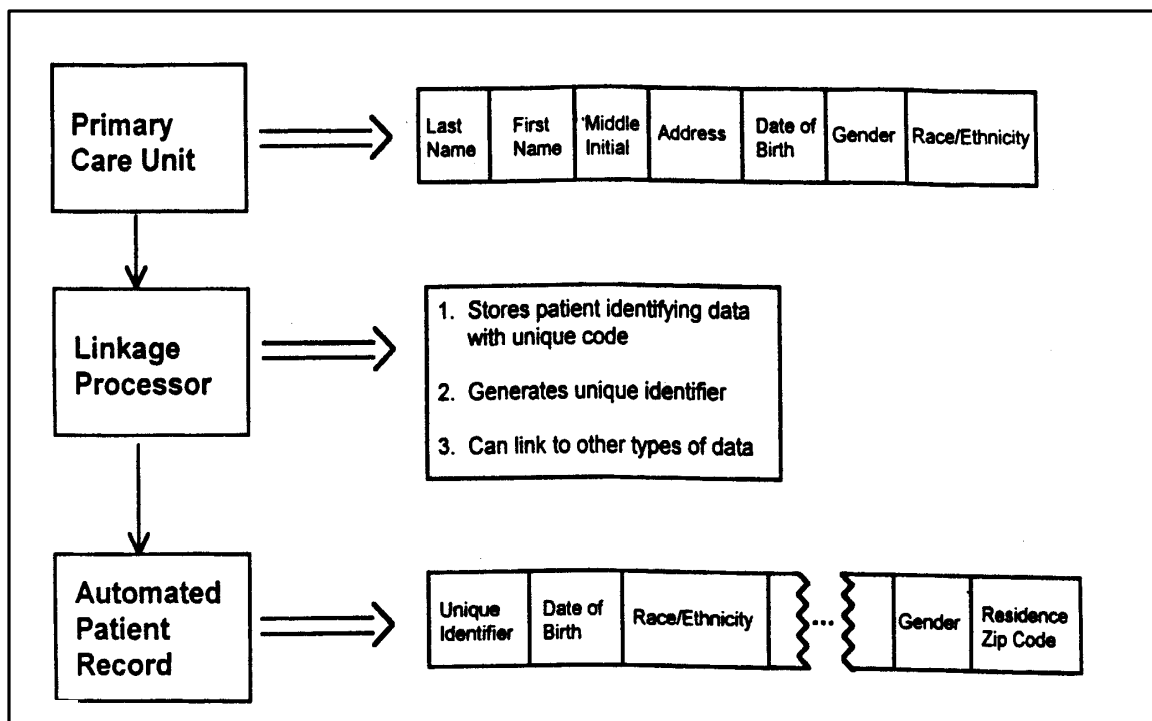


Figure 5. -- Record Linkage Architecture



2.4 Additional System Concern

In all data capture systems, of course, it is important to explicitly build-in the means to address privacy rights, the degree to which confidentiality promises are required (and kept), and the means used to make individual data physically secure. While such concerns are general, record linkage systems have some unique aspects that may bear discussion -- particularly the systems described in Section 2.3 above. Figure 6 summarizes these, emphasizing the additional complexity introduced by the linkage environment and the degree to which linkage systems are or should be "auditable." By "auditable" is meant that, at a minimum, each access to identifiable data is controlled and a log kept of the individuals who obtained the data and of all transactions that occurred (in other words, an **audit trail** is kept so that outside monitoring is possible).

Figure 6. -- Some Overall Record Linkage System Concerns

Linkage Issues	Complexity	Auditability
Privacy	Extremely high, may be beyond our current understanding, without training and experience	May be very difficult to establish, maintain, or use in monitoring access
Confidentiality		
Security		

Tore Dalenius has provided a good review of privacy, confidentiality, and security goals in statistical settings [19]. His work may afford a point of departure for the discussion here. In common speech, the words: privacy, confidentiality and security partially overlap in usage and often have meanings that depend greatly on context. Each can also have an emotional content which makes precise definitions difficult, even contentious. For example, Dalenius quotes Westin (1967) as saying about privacy:

Few values so fundamental to society as privacy have been left so undefined in social theory or have been the subject of such vague and confused writing by social scientists.

A good start on giving meaning to the word "privacy," or "information privacy" (our context here), might be the definition first articulated by Justice Brandeis as the "right to be let alone...the most comprehensive of rights and the right most valued by civilized man" [20]. Attempts to update this definition have been many and will undoubtedly continue. All afford the individual or data subject some, sometimes sole, rights over what matters they want to keep private and what matters they are willing -- or want -- to reveal.

Record linkage settings pose a particular challenge to an individual's ability to exercise his or her privacy rights. The sheer complexity of the setting makes it hard to clarify for the subject what the potential benefit or harm may be to permitting access. Consider the linkage of just two files, say, of "n" and "m" variables respectively. Cognitively for the individual involved the linkage decision may seem like one of no particular moment. The combined file will consist of data already given out earlier -- a single file of "n + m" -- rather than two separate files. But a deeper look -- at relationships, for example, between variables -- shows that a combinatorial explosion of facts about an individual has taken place -- from, say, $2^n + 2^m$ to 2^{n+m} . (Incidentally, to illustrate what this means, assume just that $n=m=11$; then the combined file has over 1000 times more information about the relationships between variables than the two files separately.)

Ready examples come to mind where individuals present themselves in one way (to get Medicaid or Medicare, say) but in another setting (perhaps a job interview) give a different, even a contradictory set of "facts."

When records from these two encounters are linked, obviously the implications may be many, since these differences would be revealed [21].

Obtaining data at different points in time and for different primary purposes is a difficulty that is peculiar to linkage settings. The privacy decision an individual may wish to make could, therefore, change over time and might depend on the particular data items as well as the purposes for which a release from their privacy rights is being sought. Singer et al., [22] for example, advocate that –

Patients sign an informed consent or a notification statement at regular intervals, not simply the first time a patient visits the provider's office.

They then go on to recommend that the consent or notification statement spell out to whom the information about the patient may be disseminated, for what purposes, and what the patient's rights are with respect to this information. Such an approach, especially as it relates to secondary uses of data and the time period for which the informed consent is valid seems clearly required in a linkage setting where patient specific information may accumulate over time and from many sources (not just one provider). It may be necessary for a regulatory body to develop regulations standardizing the contents of informed consent and disclosure policies. These regulations could definitely state what constitutes an "informed consent" and legitimate non-consented disclosure. Even then, only experience will tell whether true informed consent will be possible for most individuals.

Indeed, without wishing to jump to conclusions, it may be reasonable to conjecture that, for some kinds of data linkage at least and certain individuals, our technological abilities to electronically merge data sets may have outstripped our sense of what a data subject would have to "consent to in an **informed** way" for the systems to be built on an entirely voluntary basis. If this is so, then simply creating the health linkage system envisioned might in and of itself take away the privacy rights of some people.

The problem of complexity in record linkage systems may warrant the attention being given to complexity in general systems[23]. Linear thinking alone may, in any case, be insufficient to address what will happen not only to the individual's ability to manage his or her own data but to the system's integrity overall. What confidentiality promises can be made and kept in such a world? How can one even speculate realistically about the risks to data corruption or unauthorized disclosure? Recent experiences elsewhere [24] do not encourage belief that reasonable ways exist of being clear about even what the threats are.

Among the crucial "fail safes" is to provide an audit trail for every query against a record and any retention of a data subset. Such systems already exist for some real time settings, although not necessarily in a way that would allow a simple scale-up. A crucial step is the maintenance of these systems so they operate properly [25]. While beyond the scope of this paper, it might be noted that the expense of this maintenance step and a mechanism to "monitor the monitoring" needs a lot of attention, too. Arguments in favor of doing record linkages for efficiency reasons have not fully weighed these costs. In Brannigan and Beier [26] still other sound system architecture issues and recommendations are made that would be needed to implement essential confidentiality and security procedures, especially if large scale record linkages are to be employed.

3. Administrative Data Linkages Within the Health System

By an administrative data linkage we mean a linkage of data about an entity done with the intention of taking some direct action regarding that entity. In a health setting the most obvious example would be to assemble (link) data about a patient from multiple sources in order to improve the diagnosis or treatment of that individual. We will start with this case (subsection 3.1) and then go on to discuss administrative health linkages more generally (subsection 3.2).

3.1 Linkages for Direct Patient Care

Figure 7 lays out some of the dimensions in administrative record linkages aimed at improving the health of a patient. The Figure has five rows and two columns. Each row covers a broad area dealing with, respectively, overall issues, technical (and administrative) aspects, legal matters, the perception of the public and of experts. The primary focus of the Figure is to directly address under what conditions linkages should be permitted (Column 1) and areas for future study (Column 2). Since the goal of this paper is to just be a "starter set," only illustrative suggestions have been made in the cells, both here and elsewhere.

Among the general conditions for linkage a signed notification statement seems needed [27]. In this context, a "notification statement" might tell the patient who will have access, for what purposes and with what oversight. Hoffman in a recent paper makes the observation that "too many people may already have insufficiently monitored access to hospital patient records. He seconds Mark Siegler's thesis that "medical confidentiality, as it has been traditionally understood by patients and doctors, no longer exists." Siegler, after a patient

Figure 7. -- Administrative Data Linkages Conducted for the Health of Patients

Broad Areas	Possible Response	
	Under what conditions (Column 1)	For future study (Column 2)
Overall Recommendations	Just notification needed; if for use of patient and patient caregivers only	Concerns about coercive aspects of government "monopoly" in health care
Technical Aspects	Encryption to prevent unauthorized access and reduce risks of reidentification	Concerns about how to monitor operation
Legal Questions	For federal records, subject to the Privacy Act; use seemingly fully permitted now.	Electronic data linkages across governmental jurisdictions deserve more study; also roles of intermediaries (e.g., Health Information Trustee -- HR 4077)
Public Views	Direct evidence lacking but indirect evidence suggests that health uses to aid patients would be seen very positively.	Concerns about public view of risks associated with system need to be better understood.
Expert Opinion	An obvious use, seemingly favored by all.	Need to continue research on uniform reporting issues so as to obtain promised benefits of electronic linkages without an undue burden.

expressed concern about the confidentiality of his hospital record, scanned his medical chart and enumerated "at least 25 and possibly as many as 100 health professionals and administrative personnel...[who] had access to the patient's record," all with legitimate reasons to examine the chart [28].

Secure physical access is essential and any linkage stipulated as done for diagnosis and treatment of a patient should be available only for the use of the patient and his or her caregivers. Concerns exist about the patient data requested for such encounters and whether the demands and burdens on the patient are reasonable. The collection of uniform patient data has clear advantages; the specific data required, though, will need external review, possibly by a regulatory body -- similar to that discussed earlier on consent standards. After all there are privacy rights given up by patients to their caregivers and these should be limited to an essential minimum.

Patient and primary caregiver controlled access might involve encryption techniques or other measures designed to prevent or at least reduce the risks of unauthorized (unmonitored) use. Linkages might be time limited to reduce exposure further. As noted, Brannigen and Beier [29] have made numerous other important suggestions. System administrative issues are extensive and concerns about monitoring operations deserve continued study.

Fair information practices must be adhered to -- as required, say, in the Privacy Act and reinforced by pending legislation [30]. Continuing study of state and local restrictions [31] should be pursued to find good working models and to anticipate areas where weaknesses may arise in the National System, if litigation occurs. The *Privacy Journal* has regularly compiled state and Federal privacy laws and is a useful resource here [32].

Direct evidence of public reaction is lacking on linkages used solely to aid the patient. Such use is presumed to be very positively received. There is a large segment, though, of the public [33] that are concerned about any electronic record linkage system of the scope envisioned, mainly because of their general mistrust of government and other large institutions. These individuals or some of them, at least, might not think the benefits to be derived warrant the risks they perceive for abuse inherent in such a large-scale record linkage effort.

Virtually all "experts" take the position that notification of the use envisioned here is enough. One exception is Goldman [34] which states:

Personally-identifiable health records must be in the control of the individual. Personal information should only be disclosed with the knowing, meaningful consent of the individual.

The distinction between consent and notification may not be as important here as elsewhere. With notification there is always a "quid pro quo" -- give this data about yourself if you want to participate. In this setting patients are often asked to give what amounts to "coerced" consent; therefore, the distinction may be in name only. Logically, however, it seems inconsistent to withhold information about yourself that could be used to aid you. Unquestionably, though, a refusal to comply could mean denial of access to health care services.

3.2 Other Health Administrative Linkages

Many other health linkages are possible besides those directly involved with patient care. These could range --

- From linking treatments received by a patient to the costs of those treatments;
- To associating outcome measures (death or survival, say) to the types of medical procedures employed; and
- Even to linkages whose intent was to detect fraud or malpractice.

Data about a hospital or other health facility might be sought by looking at all the records of the patients that can be linked to that hospital. The number of possibilities, in fact, is very large -- too large to cover in any depth here. Some observations may be helpful, nonetheless, to fix a few of the ideas about what the privacy rights dimensions are:

- First, in administrative linkages such as these, the patient may become just a data point in an endeavor focussed elsewhere [35]. The dehumanizing aspect of this change of focus is inherently unsettling. Provisions like those in Figure 7 seem insufficient when the person looking at the data is not the pri-

mary caregiver but an administrator concerned about financial results, the efficiency of a medical technique, etc. -- i.e., someone without any personal relationship to the patient.

- Second, to handle the changed circumstances, among other things a "need to know" principle [36] might be applied to limit the routine availability of detailed health and demographic data. To illustrate: If data about, say, a hospital's performance is needed only hospital-level patient aggregates might be provided, rather than complete individually identifiable patient detail.

Clearly much greater safeguards seem needed once there is no longer a personal bond between the patient and the individual using the data about that patient. Arguably, establishing a convincing system that would warrant the patient and public trust required here may be exceedingly difficult.

An important issue that may deserve comment is the "final" disposition of a patient's health (and related financial) records when the patient dies. Even for federal record systems, the Privacy Act no longer offers any protection, for example. We are learning more and more about the genetic causes of some illnesses. Matching records from deceased patients could put their descendants (or other relatives) at risk for possible differential treatment. If the view is taken, as quoted above in Goldman, that the patient "owns" his or her records then, by inference, upon death the estate of the patient owns that patient's records and their disposition is a matter to be settled by the heirs. In any event, inter or intra-generational record linkage needs careful consideration and might be done, as a rule, only with the consent of **all** individuals so linked.

4. Research Data Linkages Within the Health System

It can be argued that some research uses of data linkages within the Health System are administrative and so are already covered by the discussion in Section 3, especially subsection 3.2. There can be a fine line between applied research (intending to serve a permissible administrative purpose) and basic research (involving possibly an unanticipated analysis of variables originally obtained for another purpose).

Rather than try to draw the line, however, we will confine our attention to "basic research" since this involves some potentially new issues. In particular, our discussion will focus on researchers who are in some sense outside the Health Care System -- i.e., individuals that do *not already have access* to the patient data. Such a decision has consequences, of course. For example, important issues, like what research doctors do when using data about their own patients, go undiscussed. On the other hand, there is already an extensive body of practice on this topic and record linkage issues do not seem primary.

In any event, for the basic research setting we have confined attention too, figure 8 attempts to set out a summary of the main issues. As in Figure 7 earlier, included are some overall recommendations, legal and procedural questions are addressed, as well as perceptions concerns (both by the public and among the experts). These are further elaborated below.

Notification of patients about basic research uses may be sufficient in some settings while a specific consent may be needed in others. All basic research should be authorized by a review board mechanism of some sort with an annual public report, perhaps, to an outside citizens body. Requirements for securing consent pose difficult logistical and statistical problems that need extensive study. Anonymous group matching offers a potentially promising middle ground that could allow individual consent decisions to be honored, yet may not greatly sacrifice approved scientific ends [37]. However, as Figure 8 states, an extensive development and evaluation period is needed before this approach will prove its value.

Figure 8. -- Basic Research Data Linkages within the Health System

Broad Areas	Possible Response	
	Under what conditions (Column 1)	For future study (Column 2)
Overall Recommendations	Notification and even maybe consent required for individual linkages, plus research review board authorization	Statistical properties of group linkages and their use need extensive study when consent not given.
Technical Aspects	Elimination of all obvious (and not so obvious) identifiers. Access to data also limited by reidentification risks and "need to know"	Research on use of synthetic data. Continuous study of (ever) changing reidentification risks.
Legal Questions	Laws often unreasonably require <u>no</u> risk of redisclosure.	Research on "proof of harm" issue. Legislative and litigation research on contract based research access.
Public Views	Significant negative sentiment tied to distrust of government and lack of a specific clear purpose.	Study reactions to longterm (lifelong) record linkage
Expert Opinion	For the most part strongly favor broad basic research uses requiring only notification.	Nonmedical uses of health system records need more study.

The elimination of all identifying items about a patient would seem to be a necessary prerequisite for broad access to the health system data base by outside researchers. The risks of potential reidentification [38] are an ongoing concern, especially as nonhealth electronic systems grow in size and potentially have common variables which overlap those in health data bases. Research access through contractual arrangements as proposed by Herriot [39] has already begun in some settings (where it might be evaluated) and deserves study in others (where it has yet to be applied). The development of wholly synthetic data sets [40] also warrants work

and may be potentially promising because of the public assurances that can be given which might satisfy even those who greatly distrust government.

As noted earlier, there are a significant minority of individuals who oppose linkages and this group grows larger when there is no clear and compelling purpose for such linkage, except an ill-defined one -- like "basic research." [41] Lifelong patient linkage projects which are particularly attractive basic research tools may be subject to potentially severe public reaction if done without continuing consent (as occurred in Sweden [42]).

In general, even the strongest human rights advocates make an exception for research uses of individual data, stating [43] that "Information that is not personally-identifiable may be provided for research and statistical purposes." Given the growing power of probabilistic matching, though, we may not be far from the day when the only way to remove personally-identifiable information about some individuals is to remove all direct data concerning such individuals from a research file. Additionally, there may be some concerns about the appropriateness of nonmedical uses of health care records as, say, for the decennial census, [44] a point more appropriately covered in the next section.

5. Research Linkages between Health and Other Record Systems

Our discussion of basic research issues within the health system (Section 4) forms a bridge to a discussion of research data linkages between health and other record systems. Many parallels exist, as may be seen by comparing Figure 8 with Figure 9 below. There are, however, some new elements too:

- First, deterministic matching algorithms should be possible within the health system, assuming some form of health identifier is settled on. Generally, though, unless the SSN is used as the health identifier, only probabilistic matching methods will be available between health and nonhealth record systems; hence greater uncertainty about linkage quality will exist.
- Second, these nonhealth systems were clearly intended for nonhealth purposes; thus, their use in health record linkage research, through the simple expedient of health legislation, say, seems problematic. In fact, a strong case might be made for "consent only access" to at least some of them. Also any retroactivity in this expanded use should not be taken lightly either.
- Third, there seems to be a wide range of record linkage options, spanning matches to vital records at one end of the spectrum [45] (a traditional epidemiological tool) with tax records at the other [46] (something seldom done). The views of experts and the public appear to move predictably along this continuum from some acceptance to almost none [47].
- Fourth, even anonymous group matching methods need more study in this setting and not just their statistical efficiency as noted in Figure 8 but their public acceptability. Black males seem particularly opposed to, at least, some linkages. Concerns like those in Fisher et al. [48] merit examination here too.

As already noted, at least some experts are concerned about proposals using health records to improve the accuracy of the decennial census population count [49]. In fact, except in cases where explicit consent is obtained, it may make sense to confine all matches of health records to nonhealth records solely to those research purposes related to health. The control of any linkages between health and nonhealth records, say with Census Bureau data, needs careful study too [50]. Most Federal statistical agencies, for example, currently **lack** auditable record linkage systems [51]. and would have to greatly increase internal controls to meet what should be stringent electronic access (and audit) standards [52].

Figure 9. -- Research Data Linkages between Health and Other Record Systems

Broad Areas	Possible Response	
	Under what conditions (Column 1)	For future study (Column 2)
Overall	Generally consent should be required plus research review board authorization	Same as Figure 8.
Technical Aspects	Same as Figure 8.	Same as Figure 8.
Legal Questions	Conforming legislation needed to Tax Code, Social Security Act, etc.	Research on "proof of harm" issue. Legislative and litigation research on contract-based research access.

Public Views	Significant minority would not consent to individual linkages	Research on reactions to group linkages for statistical purposes. Study parallel to HIV testing.
Expert Opinion	For the most part strongly favor health research uses only requiring notification.	Nonmedical uses of any linkages need more study.

6. Nonresearch Linkages between Health and Other Record Systems

As may be apparent by now, in this paper there has been a progression from linkage opportunities that might be viewed by most individuals as beneficial, even to be encouraged, to linkages that are more problematic. This section discusses linkages that, in the view of many, may be dangerous and should generally be discouraged.

Figure 10 sets out a summary of possible issues in nonresearch linkages between health and nonhealth systems. Some overall observations on this figure might be worth making too -- highlighting what is new or controversial.

With the exception of a court order in a criminal case, all nonresearch linkages for nonhealth reasons should be prohibited. Even health administrative linkages (say, to use IRS address information to locate a person for health reasons) should be carefully limited (as is the case now). Areas for future study might include research on notification issues and consent-based exceptions. After all, new health needs keyed to helping individuals may arise over time and hence notification statements might need to be changed or at least their understanding reviewed periodically.

Figure 10. -- Nonresearch Administrative Data Linkages between Health and Nonhealth Record Systems

Broad Areas	Possible Response	
	Under what conditions (Column 1)	For future study (Column 2)
Overall Recommendations	For nonhealth reasons only with a court order. For health reasons only to directly aid patients.	Continuing research on (changing?) understanding of all consent or notification statements.
Technical Aspects	Minimizing redisclosure risks, especially to open or decentralized systems like vital records.	Continuing research on record keeping practices in nonhealth record systems, government and private.
Legal Questions	Ban any use of a new health identifier in nonhealth record systems.	Study conforming legislative needs.
Public and Expert	In generally close agreement, with a major-	Continuous routine monitoring.

Opinion	ity favoring restrictions on nonhealth uses.	
---------	--	--

Existing systems, especially vital records, have many variables in common with health care record systems. Vital records are also quite open and hence they pose a significant risk of redisclosure, especially in public use (or other widely available) research files. If an independent health identifier is *not* used, then perhaps the SSN, for example, should be removed, or access to it restricted on birth and death records.

A legal ban, of course as generally advocated, should be imposed on the use of any new health identifier created, *except in health systems*. Research on other obvious and not so obvious identifiers, e. g., geographic details, should be ongoing to be sure that (legislated?) health record practices keep up with technology and the changing nature of unauthorized disclosure risks.

Public and expert opinion appear to both strongly oppose nonhealth administrative use of health record systems [53]. Additional public opinion research, though, seems needed on this point and others. For example, what are the public's views on the risks to any *new* health system from the *existing* centralized federal record systems (at IRS and SSA, for instance)? What about their views on the real danger of probabilistic matches to private data bases or to open or decentralized government systems, like vital records?

7. Summary Recommendations

Throughout this paper recommendations have been made that address aspects of privacy concerns in any large scale record linkage activity involving the proposed new health system or between that system and others. Figure 11 below provides a brief summary of these.

Figure 11. -- Selected Permissible Record Data Linkages by Purpose and Under What Conditions

Type of Data Linkage	Permissible and Under What Conditions
Administrative Data Linkages for the health of the patient	Just notification needed; if for use of patient and patient caregivers only
Other Administrative Data Linkages of Patient Records within the health system.	Greater safeguards seem needed once there is no longer a personal bond between patient and service provider (caregiver)
Basic Research Data Linkages within the Health System	Notification and even maybe consent required for individual linkages; research review board authorization.
Research Data Linkages between Health and Other Record systems	Generally consent should be required plus research review board authorization.
Nonresearch Administrative Data Linkages between	For nonhealth reasons, only with a court order. For

Health and Nonhealth Record Systems	health reasons, only to directly aid patients.
-------------------------------------	--

The overall treatment of linkage opportunities in this paper has gone from situations that simply called for a signed notification statement, preferably at regular intervals (Section 3), to suggested (Section 4) or required (Section 5) informed consent -- for linkage research in the health system or linked record research more generally. Finally (in Section 6), there was a brief discussion of how to **prevent** matching for nonhealth administrative purposes, except in rare instances. In all of these discussions, recommendations have been given along with the views of others; also areas for future study have been highlighted.

Frankly, this paper advocates a "go slow," careful approach to any attempt at data linkages undertaken as part of health care reform. It is unlikely that all the potential vulnerabilities of the new linkage system will be learned by anything other than experience -- hopefully not too hard won. Prototyping linkage experiments [54] are key. Patient consent and notification experiments will also be needed, as well as continuous study of public and patient opinion. An evolutionary rather than revolutionary strategy seems to represent the kind of humility and listening needed to avoid major blunders, especially in any advertent or inadvertent "takings" of privacy rights.

Much of the motivation around health reform speaks to efficiencies that can be gained with standardization of reporting and electronic data networking. These arguments seem to have merit; however, even if true, such changes will require a great many people to learn to do things in new ways and potentially paper records may need to continue to be employed for a long time (even if all new encounters are captured electronically).

Because the job is so big, it is important to begin **now but incrementally**. If structured properly, an orderly transition could be conducted, leaving ample time for human rights impacts to be respected.

8. An Afterword

An afterword may be worth making concerning the recommendations about "rights" in this paper; in particular, the rights to privacy and consent need to be set alongside the rights to universality and nondiscriminatory treatment [55].

Record linkage can aid a society in achieving advances in the well being its citizens. This point may have been lost in the detailed discussion of privacy and consent concerns. For example, the epidemiological literature is full of health studies that use record linkage techniques to advance knowledge [56].

The benefit side of record linkage can be oversold, however. A recent *Science* article may be worth quoting in this regard [57].

Over the past 50 years, epidemiologists have succeeded in identifying the more conspicuous determinants of noninfectious diseases -- smoking, for instance, which can increase the risk of developing lung cancer by as much as 3000%. Now they are left to search for subtler links between diseases and environment causes or lifestyles. And that leads to the Catch-22 of modern epidemiology. On the one hand, these subtle risks--say, the 30% increase in the risk of breast cancer from alcohol consumption that some studies suggest -- may affect such a large segment of the population that they have potentially huge impacts on public health. On the other, many epidemiologists concede that their studies are so plagued with biases, uncertainties, and methodological weaknesses that they may be inherently incapable of accurately discerning such weak associations. As Michael Thun, the director of analytic epidemiology for the American Cancer Society, puts it, "With epidemiology you can tell a little thing from a big thing. What's very hard to do is to tell a little thing from nothing

at all." Agrees Ken Rothman, editor of the journal Epidemiology: "We're pushing the edge of what can be done with epidemiology." With epidemiology stretched to its limits or beyond, says Dimitios Trichopoulos, head of the epidemiology department at the Harvard School of Public Health, studies will inevitably generate false positive and false negative results "with disturbing frequency."

Where does all of this leave things? The claim that the present paper is just a "starter set" is believed mainly to be true; but, in some places, even that may exceed current knowledge. What, in fact, many of the recommendations call for is simply more empirical work and hard thinking. Particularly crucial are two of these:

- Establishing ongoing programs of experimentation (e.g., on consent and notification statements), plus public opinion research on privacy issues, both in general and with a particular focus on record linkage [58].
- Instituting statistical work on group matching or other techniques that would lessen the tradeoff between the competing values of furthering scientific research **and** safeguarding personal privacy [59].

In the end, of course, the recommendations made here are simply the author's weighing of the evidence from the perspective of nearly 25 years of experience working on record link

Footnotes

-
- [1] Health Security Act (1993). Washington, DC: U.S. Government Printing Office. See also, for example, Donaldson, M. S. and Lohr, K. N., (eds.) (1994). *Health Data in the Information Age: Use Disclosure and Privacy*, Committee on Regional Health Data Networks, Institute of Medicine: National Academy Press.
- [2] Herriot, R. and Scheuren, F. (1975). The Role of the Social Security Number in Matching Administrative and Survey Records, *Studies from Interagency Linkages*, U. S. Social Security Administration.
- [3] Despite early advocates, like Dunn, H. L. (1946). Record Linkage, *American Journal of Public Health*, 36, 1412-1416.
- [4] Patterson, J. E. and Bilgrad, R. (1985). The National Death Index Experience: 1981-1985, *Record Linkage Techniques -- 1985*, Proceedings of the Workshop on Exact Matching Methodologies, Arlington Va.; Washington, DC: U. S. Department of Treasury.
- [5] Tepping, B. (1968). A Model for Optimum Linkage of Records, *Journal of the American Statistical Association*, 63, 1321-1332.
- [6] Fellegi, I. P. and Sunter, A. (1969). A Theory of Record Linkage, *Journal of the American Statistical Association*, 64, 1183 - 1210.
- [7] Newcombe, H. B. (1967). Record Linking: The Design of Efficient Systems for Linking Records into Individual and Family Histories, *American Journal of Human Genetics*, 19, 335-359. Newcombe, H. B.; Kennedy, J. M.; Axford, S. J.; and James, A. P. (1959), Automatic Linkage of Vital Records, *Science*, 130, 3381, 954-959. Newcombe, H. B. and Kennedy, J. M. (1962), Record Linking: Making Maximum Use of the Discriminating Power of Identifying Information, *Communications of the Association for Computing Machinery*, 5, 563-566.
- [8] See, for example, Beebe, G. W. (1985). Why are Epidemiologists Interested in Matching Algorithms? *Record Linkage Techniques*, Proceedings of the Workshop on Exact Matching Methodologies, Arlington, Va.; Washington, DC: U.S. Department of Treasury. See also, [56] and [57].
- [9] See, for example, Winkler, W. and Thibaudeau, Y. (1991). An Application of the Fellegi-Sunter Model of Record Linkage to the 1990 U. S. Census, *Statistical Division Report Series, CENSUS/SRD/RR - 91/09*. See also, Belin, T. and Rubin, D. (1995). A Method of Calibrating False-Match Rates in Record Linkages, *Journal of the American Statistical Association*, 90, 694 - 707.
- [10] Rogot, E.; Sorlie, P. D.; Johnson, N. J.; Glover, C. S.; and Treasure, D. W. (1988). *A Mortality Study of One Million Persons: First Data Book*, NIH Publication No. 88-2896, Bethesda, MD: Public Health Service, National Institute of Health.
- [11] Oh, H. L. and Scheuren, F. (1975). Fiddling Around with Matches and Nonmatches, *Proceedings of Social Statistics Section, American Statistics Association*. Also, Scheuren, F. and Winkler, W. E. (1997), Regression Analysis of Data Files that Are Computer Matched -- Parts I and II, in this volume: *Record Linkage Techniques -- 1997*, Washington, DC: National Academy Press. (Part I appeared previously in *Survey Methodology*, (1993), 19 (1) 39-58, Statistics Canada; Part II was delivered at the XII Methodology Symposium, Ottawa Canada, November 1, 1995, under the title Linking Data to Create Information and will be included in a forthcoming issue of *Survey Methodology*.)
- [12] Fair, M. (1995). An Overview of Record Linkage in Canada, presented at the American Statistical Asso-
-

- ciation Annual Meetings in Orlando, FL, August 1995.
- [13] Davis, K. (1995). Guarding Your Financial Privacy, *Kiplinger's Personal Finance Magazine*, 49.
- [14] Office of Technology Assessment (1993). *Protecting Privacy in Computerized Medical Information*, Washington, DC: U.S. Government Printing Office.
- [15] Scheuren, F. (1993). Correspondence with Dr. Elmer Gabrieli on a health identification number, in *Guide for Unique Healthcare Identifier Model*, ASTM, Philadelphia, May, 1993 draft. Ironically, public opinion poll data suggest that the American people favor the adaptation of the SSN, rather than the introduction of a new health identifier. See [33] for details.
- [16] *Health Security Act* (1993). Washington, DC: U.S. Government Printing Office.
- [17] Donaldson, M. S. and Lohr, K. N., (eds.)(1994). *Health Data in the Information Age, Use, Disclosure, and Privacy*, Committee on Regional Health Data Networks, Institute of Medicine: National Academy Press.
- [18] Schwartz, H.; Kunitz, S.; and Kozloff, R. (1995). Building Data Research Resources From Existing Data Sets: A Model for Integrating Patient Data to Form a Core Data Set, presented at the American Statistical Association Annual Meetings in Orlando, FL, August 1995.
- [19] Dalenius, T. (1988). *Controlling Invasion of Privacy in Surveys*, Continuing Education Series, Statistics Sweden.
- [20] *Olmstead v. United States*. 277 U.S. 438. 478 (1928) (Justice Brandeis dissenting).
- [21] Some implications are obvious. For example, "information in medical records can conceivably affect you for the rest of your life if revealed to an employer or insurance company," (*The Washington Post* Health Section, February 8, 1994). The obvious cases are not the only ones to be worried about, though. The combinatorial possibilities are so great that they may not only impair full consent to linkage by patients but also access decisions by data stewards.
- [22] Singer, E.; Shapiro, R.; and Jacobs, L (1995). Privacy of Health Care Data: What Does the Public Know? How Much Do They Care? Paper submitted with support from the American Association for the Advancement of Science, Science and Human Rights Program.
- [23] Horgan, J. (1995). From Complexity to Perplexity, *Scientific American*, June 1995, 104-109. See also, Waldrop, M. M. (1992), *Complexity*. New York: Simon and Schuster.
- [24] Superhack, *Scientific American*, July Issue, 1994, 17. This is a story of a group of about 600 computer "hacks," collaborating over the internet, who broke a computer security encryption algorithm. About 17 years earlier, it was predicted that this feat would take 40 quadrillion years. Once the effort started, it took 8 months! For more on this, see also, *Science*, May, 1994, 776-777.
- [25] In contrast, consider *The Washington Post*, July 18, 1994, where there is a story about how, despite an existing monitoring system, inadequate controls were used for access to sensitive information.
- [26] Brannigan and Beier (1995), Medical Data Protection and Privacy in the United States: Theory and Reality, paper submitted with support from American Association of the Advancement of Science, Science and Human Rights Program.
- [27] Singer, E., Shapiro, R., and Jacobs, L (1995), *op. Cit.*

-
- [28] Hoffman, B. (1990). Patient Confidentiality and Access to Medical Records: A Physician's Perspective, *Health Law in Canada*, 10:210-12. Siegler, M. (1982), Confidentiality in Medicine -- A Decrepit Concept, *New England Journal of Medicine*, 307:1518-21, as summarized by Cummings, N. (1993), Patient Confidentiality, *Second Opinion*, 112-116.
 - [29] Brannigan and Beier (1995), *op. cit.*
 - [30] Introduced by Condit as HR 4077 in the 103rd Congress; also reintroduced (again by Condit) in the 104th Congress as HR 435.
 - [31] As recommended by OTA (1993), *op. cit.*
 - [32] For example, Smith, R.E. (1992). Compilation of State and Federal Privacy Laws, *Privacy Journal*.
 - [33] Inferred from Harris-Equifax (1993), *Health Care Information Privacy: A Survey of the Public and Leaders*, New York: Louis Harris and Associates. See also, Blair, J. (1995), Ancillary Uses of Government Administrative Data on Individuals: Public Perceptions and Attitudes, Unpublished Working Paper, Committee on National Statistics, National Academy of Sciences. As Blair points out (and this author confirmed by calling Harris and Associates), the Harris-Equifax survey has important limitations on its interpretability; nonetheless, its main conclusions are in essential agreement with other research on privacy concerns. Blair summarizes these as well. Roughly, almost no matter how you ask the question, there are always about one sixth to one fifth of the population who oppose electronic record linkages on privacy grounds. Conversely, again almost no matter how you ask the question, about the same fraction will favor beneficial sounding linkages on efficiency grounds. The two thirds or so in the middle will differ in their opinions depending on the specifics. See also, [55].
 - [34] Goldman, J. (1994). Regarding H.R. 3137: Data needs and related issues for implementing health care reform, Statement before the House Post Office and Civil Service Subcommittee on Census, Statistics and Postal Personnel, Washington, DC. For an excellent expression of an alternative view, see Newcombe (1995), When Privacy Threatens Public Health, *Canadian Journal of Public Health*, 86, 188-192.
 - [35] Kluge, E. H. (1993). Advanced Patient Records: Some Ethical and Legal Considerations Touching Medical Information Space, *Methods of Information in Medicine*, 95-103.
 - [36] Brannigan and Beier (1995), *op. cit.*
 - [37] Spruill, N. and Gastwirth, J. (1982). On the Estimation of the Correlation Coefficient from Grouped Data, *Journal of the American Statistical Association*, 77, 614-620. Gastwirth, J., and Johnson, W.O. (1994), Screening With Cost-Effective Quality Control: Potential applications to HIV and Drug Testing, *Journal of the American Statistical Association*, 89, 972-981. Contrast Gastwirth, J. (1986), Ethical Issues in Access to and Linkage of Data Collected by Government Agencies, *Proceedings of the American Statistical Association, Social Statistics Section*, 6-13.
 - [38] See, for example, Jabine, T.B. and Scheuren, F. (1985). Goals for Statistical Uses of Administrative Records: The Next Ten Years, *Journal of Business and Economic Statistics*.
 - [39] Wright, D. and Ahmed, S. (1990). Implementing NCES's New Confidentiality Protections, American Statistical Association, *1990 Proceedings on the Section on Survey Research Methods*, Alexandria, Va.: American Statistical Association.
 - [40] Rubin, D. B. (1993). Comments on Confidentiality, A Proposal for Satisfying all Confidentiality Constraints through the Use of Multiple-Imputed Synthetic Microdata, *Journal of Official Statistics*.
 - [41] Harris-Equifax (1993) and Blair (1995), *op. cit.*, See also, [33]. Clearly, though, we do not know enough to be sure.
 - [42] Dalenius, T. (1988). *op. cit.*
-

- [43] Goldman, J. (1994). *op. cit.*
- [44] Singer, E.; Shapiro, R.; and Jacobs, L (1995). *op. cit.*
- [45] Fair, M. (1995). *op. cit.*
- [46] But see, for example, Scheuren, F. (1994). Historical Perspectives on the Estate Multiplier Technique, *Statistics of Income, Estate Tax Wealth Compendium*, U. S. Internal Revenue Service.
- [47] Scheuren, F. (1985). Methodological Issues in Linkage of Multiple Data Bases, *Record Linkage Techniques -- 1985*, Washington, DC: Department of the Treasury, Internal Revenue Service, 155-178. Scheuren, F. (1995), Review of Private Lives and Public Policy, *Journal of the American Statistical Association*, March 1995 Issue.
- [48] Fisher, J. et al. (1995). Gaining Respondent Participation: Issues of Trust, Honesty and Reliability, Paper submitted with support from American Association of the Advancement of Science, Science and Human Rights Program.
- [49] Singer, E.; Shapiro, R.; and Jacobs, L (1995). *op. cit.*
- [50] One joint control option that may be of interest arose in the project described in Rogot, E. et al. (1988). *op. cit.*
- [51] Scheuren, F. (1995). *op. cit.*
- [52] Brannigan and Beier (1995). *op cit.*
- [53] This might be inferred from the 1993 Harris-Equifax Questions on access to patient health data by insurance companies and employers. Harris-Equifax, *op. cit.* Also, from Blair (1995) and the other research started by Scheuren (1985). See [33] and [47].
- [54] Schwartz, H. et al. (1995). *op. cit.*
- [55] As elaborated in Chapman, Audrey R. (1997). Introduction: Developing Health Information Systems Consistent with Human Rights Criteria, *Health Care and Information Ethics: Protecting Fundamental Human Rights*, Kansas City, MO: Sheed and Ward, 3-30.
- [56] Cited earlier were Beebe [8] and Fair [12], among others. See also, endnotes [7], [10], and [34]. Also of note in this context is the paper by Sugarman, Jonathan, et al. (1997). Improving Health Data among American Indians and Alaska Natives: An Approach from the Pacific Northwest, *Health Care and Information Ethics: Protecting Fundamental Human Rights*, Kansas City, MO: Sheed and Ward, 88-113.
- [57] Taubes, G. (1995). Epidemiology Faces its Limits, *Science*, July 14, 1995, 164-169.
- [58] As advocated in Scheuren, F. (1985). Methodological Issues in Linkage of Multiple Data Bases, *Record Linkage Techniques -- 1985*, Internal Revenue Service and as pursued by him over the past 10 years through the sponsorship of numerous public opinion polls, asking various questions about linkage. Most of these are discussed in Blair, J. (1994). *Ancillary Uses of Government Administrative Data*, College Park, MD: University of Maryland Survey Research Center. Work at the Bureau of Labor Statistics, with focus groups and other cognitive research techniques, has also been sponsored. At this point, the summary given already in endnote [33] represents the limited state of knowledge.
- [59] Certainly, the seminal work of Spruill, Nancy and Gastwirth, Joseph (1982). On the Estimation of the Correlation Coefficient from Grouped Data, *Journal of the American Statistical Association*, 77, 614-620.

Additional References

- Acheson, E.D. (1967). *Medical Record Linkage*, Oxford, U.K.: Oxford University Press.
- Copas, J.B. and Hilton, F.J. (1990). Record Linkage: Statistical Models for Matching Computer Records, *Journal of the Royal Statistical Society, Ser. A*, 153 (Part 3), 287-320.
- Dunn, H.L. (1946). Record Linkage, *American Journal of Public Health*, 36, 1412-1416.
- Jaro, M. A. (1989). Advances in Record-Linking Methodology as Applied to Matching the 1985 Census of Tampa, Florida, *Journal of the American Statistical Association*, 84, 414-420.
- Kilss, B. and Alvey, W. (eds.) (1985). *Record Linkage Techniques -- 1985, Proceedings of the Workshop on Exact Matching Methodologies*, Arlington, Virginia, May 9-10, 1985), Washington, DC: Department of the Treasury, Internal Revenue Service.
- Newcombe, H.B. (1967). Record Linking: The Design of Efficient Systems for Linking Records into Individual and Family Histories, *American Journal of Human Genetics*, 19. 335-359.
- Newcombe, H.B., and Kennedy, J.M. (1962). Record Linking: Making Maximum Use of the Discriminating Power of Identifying Information, *Communications of the Association for Computing Machinery*, 5, 563-566.
- Rogot, E.; Sorlie, P.D.; Johnson, N.J.; Glover, C.S.; and Treasure, D.W. (1988). *A Mortality Study of One Million Persons: First Data Book*, NIH Publication No. 88-2896, Bethesda, MD: Public Health Service, National Institute of Health.
- Roos, L.L.; Wajda, A.; and Nicol, J.P. (1986). The Art and Science of Record Linkage: Methods that Work with Few Identifiers, *Computers in Biology and Medicine*, 16, 45-57.
- Scheuren, F. (1985). Methodological Issues in Linkage of Multiple Data Bases, *Record Linkage Techniques -- 1985*, Washington, DC: Department of the Treasury, Internal Revenue Service, pp. 155-178.
- Scheuren, F.; Alvey, W.; and Kilss, B. (1986). Record Linkage for Statistical Purposes in the United States, *Proceedings of the Workshop in Computerized Record Linkage in Health Research*, held in Ottawa, Ontario, May 21 - 23, 1986, G.R. Howe and R.A. Spasoff, (Eds.), Toronto: University of Toronto Press, pp. 198-210.
- Donaldson, M.S. and Lohr, K.N., (Eds.)(1994). *Health Data in the Information Age: Use, Disclosure and Privacy*, Committee on Regional Health Data Networks, Institute of Medicine: National Academy Press.

Note: This paper was commissioned for the Health Care and Information Ethics project, sponsored by the American Association for the Advancement of Science (AAAS) when Fritz Scheuren was working for George Washington University. It appeared as a chapter in Audrey R. Chapman (Ed.), (1997). *Health Care and Information Ethics: Protecting Fundamental Human Rights*, Kansas City, MO: Sheed and Ward. The paper is reprinted here with permission. c 1997 by American Association for the Advancement of Science. All rights reserved.

Except as permitted under the Copyright Act of 1976, no part of this paper may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by an information storage and retrieval system without permission in writing from the Publisher. Sheed & Ward is a service of The National Catholic Reporter Publishing Company. To order, write Sheed & Ward, 115 E. Armour Blvd., PO Box 419492, Kansas City, MO, 64141-6492; or call (800) 333-7373.